IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 6 February 2024, accepted 22 February 2024, date of publication 26 February 2024, date of current version 6 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3370604

RESEARCH ARTICLE

Pavement Defect Detection Algorithm Based on Improved YOLOv7 Complex Background

ZOU CHUNLONG, HUANG PEILE, WANG SHENGHUAI, WANG CHEN, AND WANG HONGXIA
School of Mechanical Engineering, Hubei University of Automotive Technology, Shiyan 442002, China

Corresponding author: Wang Shenghuai (20090023@huat.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51675167, in part by the Key Research and Development Project of Hubei Province of China under Grant 2021BAA056, in part by the Natural Science Foundation of Hubei Province of China under Grant 2020CFB755, in part by the Research Project of Education Department of Hubei Province of China under Grant T2020018 and Grant Q20191801, and in part by Hubei Provincial Outstanding Colleges and Universities Young and Middle-Aged Science and Technology Innovation Team Project under Grant T2022027.

ABSTRACT The detection of pavement diseases is an important and basic link in the road maintenance process. Many methods based on deep learning have been applied. However, these methods are not accurate enough and cannot accurately identify defects in complex background with shadow occlusion and uneven lighting brightness. In order to overcome the shortcomings of previous detection methods, a complex background defect detection algorithm based on improved YOLOv7 is proposed. First, the K-means++ clustering algorithm is used for initial anchor box setting to obtain better anchor box parameters; then, the group spatial pyramid pooling module SPPCSPC_G is introduced to replace the original SPPCSPC module to improve the fusion speed of image features and thereby improve the detection accuracy; Finally, the GELU activation function is used as the activation function of the REPConv convolution module in the YOLOv7 model, which effectively reduces model overfitting and thereby improves model detection accuracy. The test results show that the average accuracy of the improved detection algorithm for disease detection increased from 65.4% to 72.3%, an increase of 6.9%, the amount of calculation and parameters decreased by 4% and 14.9% respectively, and the FPS reached 80, an increase of 17%, and no pavement defects are missed or wrongly detected. It is more suitable for real-time detection of defects in complex background. It can be seen that the improved YOLOv7 has better detection effect on complex background defects.

INDEX TERMS Complex background, defect detection, YOLOv7, K-means++, grouped spatial pyramid pooling module, activation function.

I. INTRODUCTION

With the continuous development of our country's society, the country's investment in highways continues to increase. As of the end of 2021, the total mileage of our country's road network has reached 5.2807 million km, and the expressway has reached 169,100 km [1], ranking among the top in the world. The dense road network greatly facilitates information exchange and resource allocation between regions, thereby promoting rapid economic and social development. At the same time, the rapid development of highways has also led to a rapid increase in the number of cars and an increase in the frequency of highway use, eventually leading to the occurrence of various diseases on the highways. These diseases will affect the life of the road surface and even cause traffic accidents in severe cases. Therefore, various traffic safety management department need to conduct scientific inspections of road surface regularly to detect diseases in a timely manner.

Traditional manual visual inspection of pavement diseases has a complex process, low detection efficiency, and is accompanied by a large amount of labor costs. With the advancement of photography technology and deep learning, technical support has been provided for the collection and recognition of pavement disease images. Although many methods based on deep learning have been applied, these

The associate editor coordinating the review of this manuscript and approving it for publication was Yizhang Jiang.

methods are not accurate enough and cannot accurately identify defects in complex background with shadow occlusion and different light brightness.

In response to the above problems, scholars specializing in road defect identification and detection have conducted a lot of research and proposed many new algorithms. In terms of traditional algorithms, Su Xiaobo [2] used MobileNetv2 as the backbone network and replaced other ordinary convolutions with depth-separable convolutions. Although the efficiency of model detection has been improved, the accuracy of recognition by this algorithm needs to be improved. Zhang Zhihua et al. [3] introduced the residual neural network (ResNet) as an encoder to solve the problem of difficulty in accurately distinguishing asphalt pavement diseases with similar characteristics such as cracks and potting cracks. However, the diseases detected by this algorithm are too single, and its generalization ability to multiple diseases on highway pavement is poor. Li Erwei [4] achieved refined scanning of a given area by conducting three-dimensional scanning and image collection of bridges, but this model will be interfered by other textures in the image, affecting the accuracy of the prediction results. Liu Xing et al. [5] proposed a target detection algorithm that combines PeleeNet and YOLOv3 to improve the stability and detection rate of surface crack detection on low computing power computing platforms. He Tiejun et al. [6] proposed an improved detection model Pavement Damage-YOLO (PD-YOLO) based on pavement disease characteristics, which further improved the accuracy of pavement disease detection. In order to achieve rapid and intelligent detection of road defects. Chen Jianyu et al. [7] proposed a YOLOv5-AC algorithm. Zuo Hao et al. [8] proposed a road defect detection algorithm based on improved YOLOv5, and introduced the CBAM attention mechanism to solve the problem of detection. Inefficiency problem. However, the above scholars have made improvements based on traditional algorithms. The traditional algorithms have low accuracy and are easily affected by environmental noise, resulting in poor recognition performance and inaccurate target edges. The detected target diseases tend to be similar to those under normal light. These algorithms do not solve the problem of low target recognition accuracy under the influence of shadow occlusion and light brightness.

Currently, deep learning technology has been widely used in road surface defect detection. As a typical single-stage target detection algorithm, YOLO is widely used in system real-time detection due to its fast running speed. The detection effect of YOLOv7 on public road data sets shows that its accuracy and speed exceed other YOLO series models, so this article chooses YOLOv7 is used as an algorithm model for defect detection in complex background. Many scholars use the YOLOv7 algorithm [9], [10], [11], [12] for target detection and have achieved remarkable results. Huang P et al. [13] proposed a lightweight road defect detection model based on the improved YOLOv7 architecture. The model introduces four key enhancements that successfully balance the amount of parameters and computation, making it a more suitable algorithm for pavement defect detection. Ni Changshuang et al. [14] targeted the imaging characteristics of laser images and used a combined filter-three-histogram equalization algorithm, adding a multi-head self-attention mechanism and the funnel activation function F-ReLU to improve training accuracy; Liu P et al. [15] proposed A yellow peach target detection model (YOLOv7-Peach) based on improved YOLOv7. The K-means clustering algorithm is used to update the anchor frame information of the original YOLOv7 model, the CA (Coordinate Attention) module is embedded into the YOLOv7 backbone network, and better performance is achieved under different weather conditions by replacing the object detection regression loss function with EIoU. Effect. Chen J et al. [16] proposed a recognition method based on the improved YOLOv7 model, embedding (SENet) network and coordinate attention (CA) in the last layer of the backbone network and head network respectively, using SIoU bounding box regression The loss function refines regression inference bias and improves bounding box prediction accuracy. Zhang Z et al. [17] proposed an improved YOLOv7 traffic sign detection method CN-YOLOv7, which introduced comprehensive attention to ConvNeXt_block in the feature extraction network, and replaced the SPPCSPC module with the SPPFCPC module to reduce the number of network parameters. The K-means++ clustering algorithm is used to obtain predefined anchor frames, making them more suitable for the data set and improving the robustness and detection accuracy of the algorithm. Yolov7 exceeds all currently known detection algorithms, both in terms of speed and accuracy. However, it performs poorly in defect detection in complex background involving shadow occlusion, object overlap, and insufficient lighting. Therefore, it is necessary to modify the original model framework to achieve accurate identification of pavement defects.

In summary, in order to overcome the shortcomings of previous detection methods and better identify road defects in complex background, this paper proposes a complex background defect detection algorithm based on improved YOLOv7.

1) Use K-means++ clustering algorithm to optimize the initial anchor box to obtain better anchor box parameters;

2) Introduce the group spatial pyramid pooling module SPPCSPC_G to replace the original SPPCSPC module to improve the fusion speed of image features and thereby improve the detection accuracy;

3) Finally, the GELU activation function is used as the activation function of the REPConv convolution module in the YOLOv7 model, which effectively reduces model overfitting and thereby improves the model's detection accuracy for complex background defects.

II. YOLOv7 NETWORK MODEL

As a typical single-stage target detection algorithm, the YOLO series is widely used in system real-time detection

due to its fast running speed. YOLOv7 [18] has excellent detection accuracy and detection speed in the YOLO series model, so YOLOv7 is selected as the asphalt pavement in this article. Algorithmic model for disease detection.

The YOLOv7 algorithm introduces strategies such as Extended-ELAN (Extended-Efficient layer aggregation network) and model scaling and convolution re-parameterization based on Concatenation-Based models., which can deepen the depth of the network and improve the accuracy of the network. The model consists of the input end, the backbone architecture and the head architecture, as shown in Figure 1.

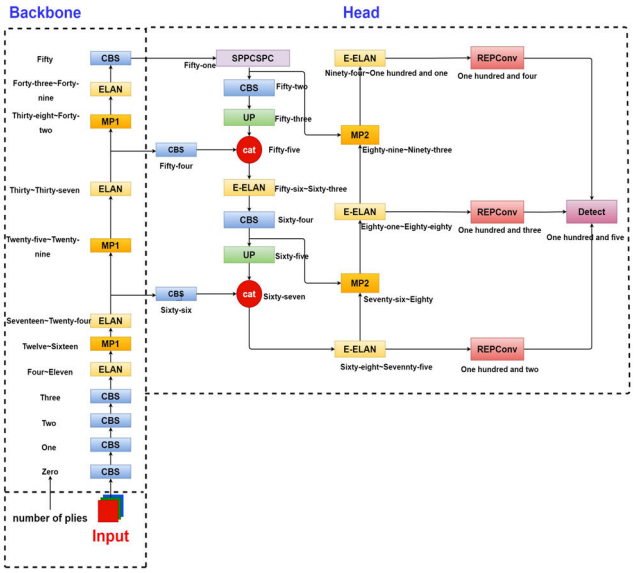


FIGURE 1. YOLOv7 structure diagram.

First, the input end is the input layer. Its main function is to scale the input image to the same size and then input it into the backbone architecture to meet the training requirements of the backbone network; the backbone architecture is also called the feature extraction layer, which consists of 50 layers (Layer0∼50) It is composed of different convolution combination modules. The main function is to extract three different sizes of target information features and input them into the head architecture. The output positions of the backbone architecture are located at the 24th layer, the 37th layer and the 50th layer respectively; the head architecture mainly includes SPPCSPC layer (Spatial Pyramid Pooling Connected Spatial Pyramid Convolution, Spatial Pyramid Pooling Connected Spatial Pyramid Convolution), E-ELAN layer, several Conv layers (Convolution, convolution layer), MP layer (MaxPool, maximum pooling layer) and REPConv Its biggest feature is the adoption of the efficient E-ELAN network architecture. ELAN-A increases the number of channels from 4 times to 8 times, that is, high channels have stronger feature expression capabilities. E-ELAN does not use the summation method of residuals, but adopts the stacking method. There is no doubt that the calculation amount is larger, but the representation is stronger. The head architecture is on the 75th, 88th and 101st layers. Three types of feature maps of different sizes are output, and the output features of different scales are adjusted in the number of image channels through the heavy-parameterized structure REP layer, and converted into bounding boxes, categories and confidence information. The convolutional layer is then used as the detection head for down-sampling to achieve multi-scale detection of large, medium and small targets. Figure 2 shows the structure diagram of each module.

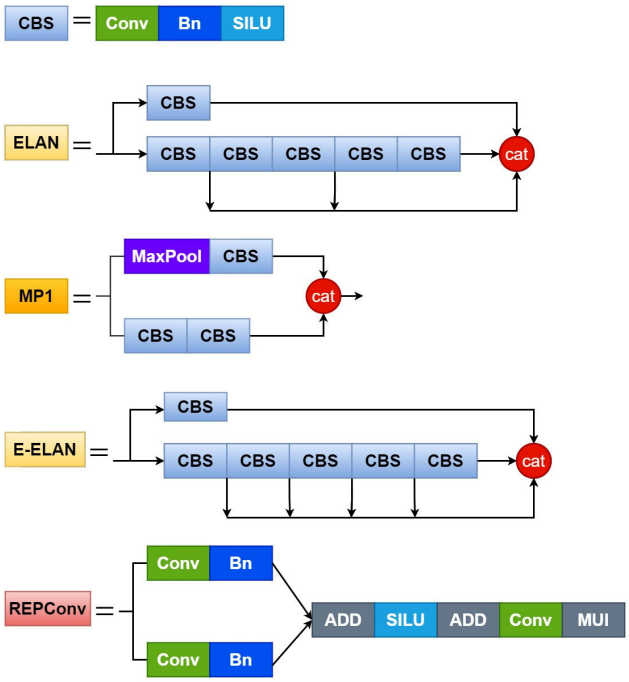


FIGURE 2. Structure diagram of each module.

III. IMPROVED YOLOv7 NETWORK MODEL

A. K-MEANS++ CLUSTERING PRIOR BOX

The implementation of the YOLO series of algorithms requires traversing the preset pixel frames in the image, retaining the best pixel frames and making fine adjustments. The above preset pixel frames are called anchor frames. YOLOv7 uses the K-Means algorithm by default to cluster the anchor frames generated by the COCO data set. Since the COCO data set has many categories and large differences in target sizes, if the initial Anchor is used for training, the convergence speed of the model will slow down and the model will In the end, the detection effect of the model is poor, so the Anchor clustered by the K-means algorithm is not suitable for the data set of this article. In addition, when K-means is clustering, the convergence situation depends heavily on the initialization status of the cluster center, which may lead to a large gap between the randomly assigned initialization cluster center and the optimal cluster center. Therefore, this article uses the K-means++ algorithm instead of the original K-means algorithm to select better clustering centers, better identify road defects under complex

backgrounds such as uneven brightness, and reduce false detections and misdetections. Improve detection accuracy and effect.

Through experiments, it was found that K-Means-new uses 1-IoU (Intersection over Union) to indicate that the distance calculation effect will be much better. However, when conducting experiments on the data set in the article, it was found that IoU is used to establish distance measurement indicators, and K-means++ is used. The algorithm combines the Intersection over Union (IoU) to set the anchor box. The K-means++ algorithm randomly selects the first initial cluster center when selecting it, and selects the a-th initial cluster center (1< a<K), the farther a point is from the first a-1 cluster center, the greater the probability of being selected as the a-th cluster center, and the initial anchor box size distribution is more uniform and more representative. While improving the randomness and locality of the clustering center selection process, it can also better improve the detection accuracy and effect, thereby improving the judgment of defects by the attention mechanism in the next step and suppressing irrelevant background areas.

The specific steps of the K-means++ anchor box setting scheme: first traverse each target in each label in the FLIR data set, randomly select a target as the initial cluster center, and divide it into 9 regions, and calculate each target in the data set. The distance between the target and the initialized cluster center, and the target with the largest distance is selected as the new cluster center with probability. At the same time, the probability of each target being selected as the next cluster center is calculated, and the target with the largest probability is selected as the next cluster. center, its calculation formula is as follows:

Step 1: Randomly select a center point m_i among the data points,

Step 2: Use the IoU distance square to calculate the distance $D(x)$ between the remaining sampling points x and the center point m_i,

$$D_{(x)} = \|x - m_i\|_2^2 \tag{1}$$

Step 3: Calculate the probability $P(x)$ of each sampling point becoming the new cluster center point. Select the point with the largest probability value to become the new cluster center,

$$P(x) = \frac{D(x)^2}{\sum_{x \in X} D(x)^2} \tag{2}$$

Repeat steps two and three until k initial clustering center points are selected. For each initial clustering center point $i \in \{1,2,3, \cdots\cdots,k\}$, define the nearest point set M and update the M set center of mass.

Although the K-means++ algorithm takes a little more time to initialize the cluster center than K-means, it will converge quickly after selecting the appropriate cluster center. Overall, it reduces the calculation time and improves the calculated Anchor and Matching accuracy of targets in the dataset. It is very effective for road target detection in complex background.

B. INTRODUCING THE SPPCSPC-G GROUPING SPATIAL PYRAMID POOLING MODULE

The SPPCSPC (Spatial Pyramid Pooling with Channel Spatial Pooling) module is a convolutional neural network module used for image classification. It can improve the receptive field and classification accuracy of the model. The following are the implementation steps of the SPPCSPC module:

1. The size of the input feature map is H/W/D, where H represents the height of the feature map, W represents the width of the feature map, and D represents the number of channels of the feature map.

2. For the channel dimension, perform operations such as maximum pooling, average pooling, and minimum pooling to obtain D pooling results, with the size of each result being 1/D.

3. For the spatial dimension, use pyramid pooling to divide the feature map into different areas, and perform operations such as maximum pooling and average pooling on each area to obtain n pooling results, where n represents the partition. quantity.

4. Splice the channel pooling results and the spatial pooling results together to obtain a feature vector.

5. For each feature vector, a fully connected layer is used for tasks such as classification or regression. Figure 3 is a comparison diagram of the structures of SPPCSPC and SPPCSPC_G.

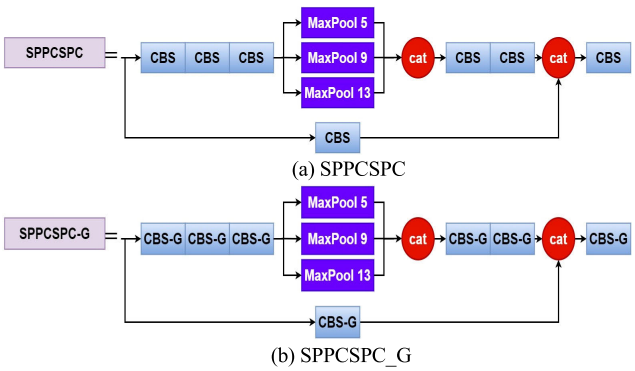
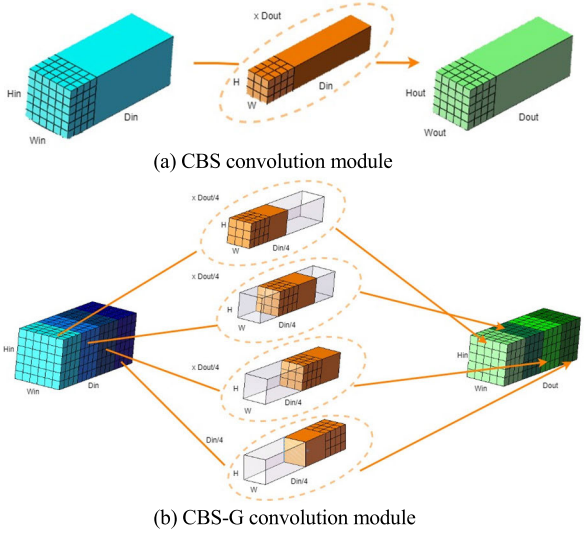


FIGURE 3. SPPCSPC and SPPCSPC_G structure comparison chart.

Aiming at the problem of pavement defect detection in complex environments in this article, the results are not good. The spatial pyramid pooling module only cares about the task-related areas. When the target area is occluded or interfered by light, it will affect the judgment of the attention mechanism. In order to solve the above problems, this paper proposes to design the grouped spatial pyramid pooling module SPPCSPC_G, which changes all seven convolution modules to perform grouped convolution in the original SPPCSPC. Each convolution module is divided into four groups, and group convolution has the following

advantages. The first advantage is efficient training. Because the convolutions are split into multiple paths, each of which can be processed separately by a different GPU, the model can be trained on multiple GPUs in parallel. The second advantage is that the model will be more efficient, that is, the model parameters will be reduced as the number of filter groupings increases. The grouping spatial pyramid pooling module will dynamically identify image features in multiple dimensions at the same time, and in this way The mechanism of emphasizing areas of interest and suppressing irrelevant background areas makes the grouped spatial pyramid pooling module more accurate than standard two-dimensional convolution in identifying target defects in complex background. The third advantage is that the grouped spatial pyramid pooling module greatly reduces the amount of parameters and calculations of the model. The improved spatial pyramid pooling module SPPCSPC_G grouping can reduce the amount of parameters and calculations while enhancing the model's focus on deep semantic features to suppress useless information. Enhance the ability to distinguish the characteristics of large and small pavement defects. Figure 4 is a comparison of the original CBS convolution module and the improved CBS-G convolution module.



(a) CBS convolution module

(b) CBS-G convolution module

FIGURE 4. Comparison chart between CBS convolution module and improved CBS-G convolution module.

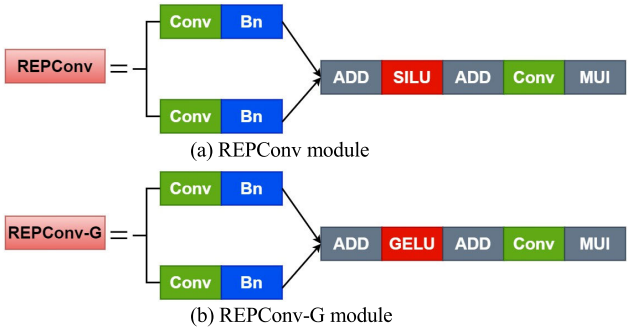
For the CBS module, we can see from the figure that it consists of a Conv layer, which is a convolution layer, a BN layer, which is a Batch normalization layer, and a Silu layer, which is an activation function. The CBS-G module here changes the number of channels of the feature map from one channel to 4 channels. Because the convolution is divided into multiple paths, each path can be processed separately by different GPUs, so the model can be processed in parallel. Training is performed on multiple GPUs, thereby reducing the number of parameters and calculations of the model.

C. REPLACING THE GELU ACTIVATION FUNCTION

Each neuron node in the neural network accepts the output value of the neuron of the previous layer as the input value of this neuron, and passes the input value to the next layer. The input layer neuron node will directly pass the input attribute value to the next layer. One layer (hidden layer or output layer). In a multi-layer neural network, there is a functional relationship between the output of the upper layer node and the input of the lower layer node. This function is called the activation function. If the feature information transfer between layers is through linear changes, the model will become easy to verify, and the approximation ability of the network will be very limited, so an activation function is required. The original YOLOv7 model network structure continues the SiLU activation function used in YOLOv5, which has the characteristics of lower bound, no upper bound, smoothness, and non-monotone. The mathematical expression of the SiLU activation function is:

$$silu(x) = x^{*}sigmoid(x) = \frac{x}{1+e^{-x}} \tag{3}$$

Because the YOLOv7 network model has too many parameters, the model complexity is too high, which can easily lead to overfitting and the model has spatial insensitivity. The GELU activation function is related to random regularization and can have the effect of self-adaptive dropout, that is, it can effectively reduce model overfitting, thereby improving model detection accuracy. And because the calculation of GELU is relatively complex, replacing too many convolution modules will consume more computing resources, thus reducing the detection speed of the model. Therefore, this study uses the GELU [23] function as the activation function of the REPConv convolution module in the YOLOv7 model. The modified model convolution module is shown in Figure 5(b).



(a) REPConv module

(b) REPConv-G module

FIGURE 5. Comparison chart between REPConv module and REPConv-G module.

GELU is an activation function based on the Gaussian error function, the full name is "Gaussian Error Linear Unit". As an excellent activation function proposed in 2020, compared with SiLU, the function will not treat all x less than or equal to 0 equally as 0., taking all 0 will cause the derivative to be equal to 0, causing the gradient to disappear, so the GELU activation function eliminates the problem

of the gradient disappearing. At x=0, the GELU function introduces a transformation similar to the sigmoid function in the nonlinear transformation of the activation function, which allows the output of the GELU function to fall within a wider range, helping to accelerate the convergence of the model. The following is the mathematical expression of the GELU activation layer:

$$GELU\ (x) = x * P\left(X \leq x\right) = x * \Phi\left(x\right) \tag{4}$$

where $\Phi\left(x\right)$ represents the cumulative distribution function of the normal distribution, that is:

$$\Phi\left(x\right) = \frac{1}{2} \cdot \left(1 + erf\left(\frac{x}{\sqrt{2}}\right)\right) \tag{5}$$

$erf\ (x)$ represents the Gaussian error function. This function can further be expressed as

$$x * P\left(X \leq x\right) = x \int_{-\infty}^{x} \frac{e \cdot -\frac{(X-\mu)^2}{2\sigma^2}}{\sqrt{2\pi}\,\sigma} dX \tag{6}$$

Compared with the SiLU function, the GELU function has a non-zero gradient in the negative value area, thereby avoiding the problem of dead neurons, improving the neural network's ability to express the model, and solving problems that cannot be solved by linear models. In addition, GELU is smoother than SiLU near 0, so it is easier to converge during the training process, thereby improving the feature extraction of road defects in complex background. By processing negative samples, the proportion of negative sample loss contributions is reduced, and positive samples are highlighted. The loss contribution is used to improve the model's detection performance of dense small target diseases in road defects.

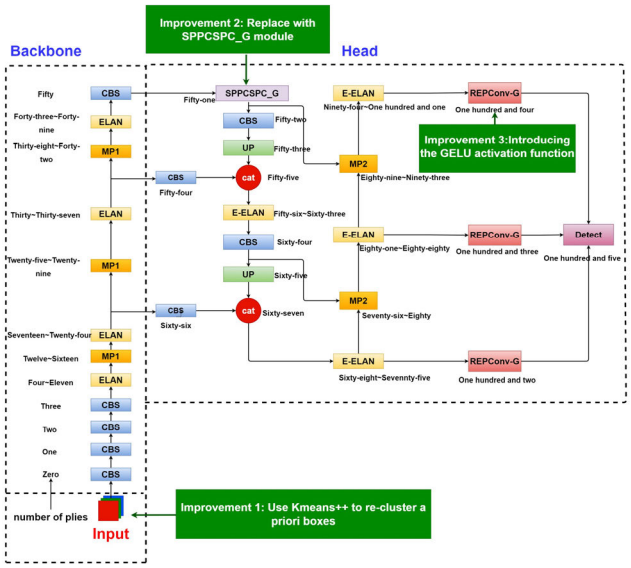


FIGURE 6. Improved YOLOv7 model.

D. IMPROVED YOLOv7 MODEL

As shown in Figure 6, it is the YOLOv7 model network structure improved in this article. First, the K-means++ clustering method is used to cluster a priori boxes suitable for this data set based on the ratio of the length and width of the annotated boxes in the defect data set, which improves the recall rate of the network, speeds up the convergence of the a priori boxes, and reduces Reduce the time of inference testing; introduce the group spatial pyramid pooling module SPPCSPC_G to replace the original SPPCSPC module to improve the fusion speed of image features, thereby improving the detection accuracy; finally, replace the activation function of the REPConv module with the GELU function to effectively reduce model overfitting, thereby improving model detection accuracy.

IV. EXPERIMENT AND RESULT ANALYSIS

The overall flow chart of the experiment is shown in Figure 7.

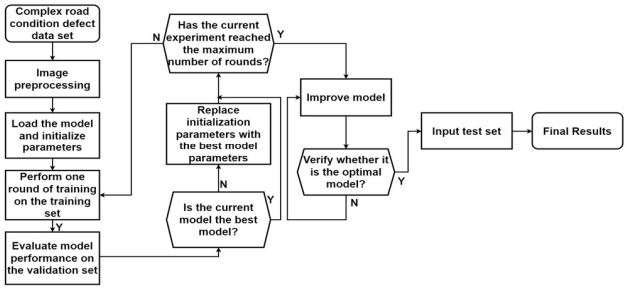


FIGURE 7. Overall flow chart of the experiment.

A. DATASET CONSTRUCTION

This experiment selected 2525 images with more disease defect characteristics from the data set (RDD-2019) [19]. Due to the lack of complex background defects in public data sets, this experiment also added a self-collected complex background defect image data set. The collection method is to fix the mobile phone on the car sun visor and take pictures of various road defects as the car moves forward. Finally, 510 photos were selected and labeled using the open source image annotation tool LabelImg. The ratio of shadow occlusion, uneven lighting and normal road defects in this paper's data set is 3:3:4. There are 3035 images in the RDD-2019 and self-made data sets. The validation set, test set and training set are divided in the ratio of 1:1:8. Table 1 is a statistical table of various types of quantities.

B. EXPERIMENTAL CONFIGURATION

This experiment uses the PyTorch training framework, uses Win10 as the system environment, the GPU model is NVIDIA V100, the video memory size is 16GB, the deep learning environment is Python 3.8, the framework is Pytorch 1.10.0, and the GPU acceleration is CUDA10.2. Table 2 shows the network model parameter settings.

Figure 8 shows an example of a sample image.

TABLE 1. Statistics table of the number of various tags.

Table	Category	Amount
D00	Longitudinal cracks	933
D10	Transverse cracks	1176
D20	Network cracks	1676
D43	Blurred sidewalk	982
D44	Blurred white lines	1722
D50	Manhole cover	1147

TABLE 2. Network model parameter settings.

Parameter settings	illustrate
Batch_size=16	The number of batches is 16
Epoch=150	Training data for 150 rounds
Learing_rate=0.01	The initial learning rate is 0.01
SGD	Optimizer
Weight_decay=0.0005	The weight attenuation coefficient is 0.0005

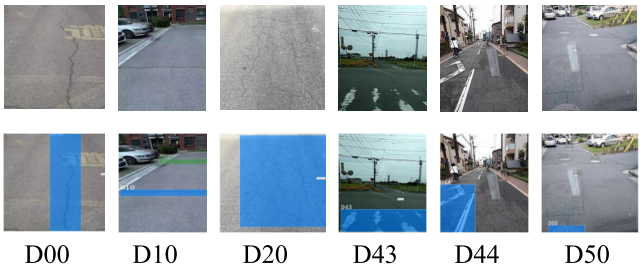


FIGURE 8. Sample image example.

C. EVALUATION INDICATORS

In order to evaluate the model's performance in identifying and detecting pavement defects, the evaluation criteria include accuracy (precision), recall (recall), mean average precision (mAP) and frames per second.

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$R = \frac{TP}{TP + FN} \tag{8}$$

$$mAP = \frac{1}{m}\sum AP(i) \tag{9}$$

$$FPS = \frac{n}{T} \tag{10}$$

In Formulas 7 to 10: TP is the number of correctly recognized detection frames, and FP is the number of incorrectly recognized detection frames. FN is the number of correct targets that have not been detected, m is the number of detected categories, AP is the area under the PR curve, which reflects the model's recognition effect on a certain category, n is the number of pictures processed by the model, and T is the consumption time.

D. EXPERIMENTAL RESULTS

The experiment set up comparative experiments on three improvement measures, namely replacing the corresponding a priori frame, introducing the SPPCSPC_G grouping space pyramid pooling module, and replacing the GELU activation function, to analyze the impact of different improvement measures on the network detection effect. Finally, in order to verify the proposed method in the article Each optimization part performs ablation experiments on the impact of the algorithm.

1) RESET THE A PRIORI FRAME COMPARISON EXPERIMENT

The anchor processing mechanism of YOLOv7 is to recalculate the anchor when the anchor in the configuration file calculates the best possible recall (BPR) less than 0.98. The maximum value of BPR is 1. If BPR is less than 0.98, the program will automatically learn the size of the anchor based on the label of the data set. The calculated initial anchor frame size is shown in Figure 9. Comparison shows that the initial anchor frame size distribution in (c) is more even and representative.

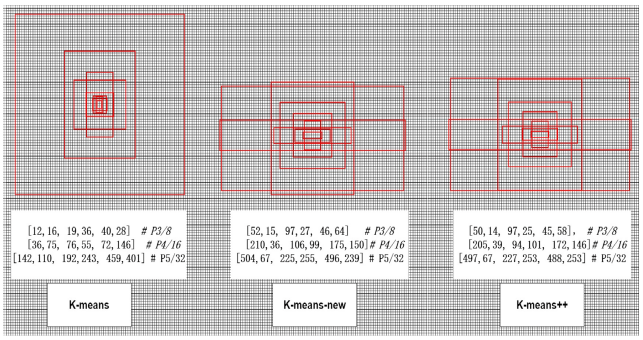


FIGURE 9. Initial anchor box clustering.

TABLE 3. The impact of resetting the a priori frame on network performance.

Priori box	mAP@0.5%	FPS	Anchor to target frame ratio	BPR
K-means	65.4	68	3.87	0.9847
K-means-new	67.4	65	4.69	0.9996
K-means++	68.6	70	4.68	0.9996

As can be seen from Table 3, it is found through experiments that the ratio of the output anchor to the matching target frame when training using the Euclidean distance

K-means algorithm is 3.87, the BPR is 0.9847, and K-means-new clustering is represented by 1-IoU as the distance. The ratio of the output anchor to the matched target frame reaches 4.69, the BPR reaches 0.9996, and the detection accuracy is slightly improved. However, the K-means++ algorithm is used in combination with the Intersection over Union (IoU) to set the output anchor and matching of the anchor frame. The target frame ratio reaches 4.68, and the BPR reaches 0.9996. When the anchor and matching target frame ratios are almost the same, the K-means++ detection accuracy is greatly improved, and the inference speed is also greatly improved, indicating the effect of K-means++ clustering. Better than the original K-means algorithm.

2) INTRODUCING THE SPPCSPC_G GROUPED SPATIAL PYRAMID POOLING MODULE

TABLE 4. Introducing SPPCSPC_G comparison experiment.

Module	Location	mAP@0.5%	FPS	Flops/G	Params/M
SPPCSPC	51	65.4	68	104.8	37.22
SPPCSPC_G	51	69	73	100.3	31.52

It can be seen from the experimental results in Table 4 that a grouped spatial pyramid pooling module SPPCSPC_G is designed in this paper. After replacing the original SPPCSPC with SPPCSPC_G, the accuracy is greatly improved, reaching 69%, and the calculation amount and parameter amount are reduced by 4.3% and 15.4% respectively. the model is more efficient, and it reduces the amount of parameters and calculations while improving detection accuracy.

3) COMPARATIVE EXPERIMENT OF REPLACING THE GELU ACTIVATION FUNCTION

TABLE 5. The impact of changing the activation function on the network.

Activation function	mAP@0.5/%	FPS	Flops/G	Params/M
SiLU	65.4	68	104.8	37.22
GELU	68.3	71	104.8	37.22

From the experiments in Table 5, it can be seen that because the YOLOv7 network model has too many parameters, the model complexity is too high, which can easily lead to overfitting and the model has spatial insensitivity problems. The GELU activation function is related to random regularization and can have the effect of self-adaptive dropout, which effectively reduces model overfitting and improves model detection accuracy. It can also be clearly seen from Table 5 that mAP has increased from 65.4% to 68.3%. improves detection accuracy and speeds up model inference.

4) ABLATION EXPERIMENT

It can be seen from the experiments in Table 6 that the first set of experiments uses the original YOLOv7 model as the basis, with a detection accuracy of 65.4%, a model calculation amount of 104.8G, and a model parameter amount of 37.22M; the second set of experiments uses K-means++ for the initial anchoring of the model. At the same time as setting the frame, the SPPCSPC_G grouped spatial pyramid pooling module was introduced to speed up the convergence of the model and improve the global feature learning ability of the model. The detection accuracy reached 69.8%, which was 4.4% higher than the original model. The amount of calculation and parameters They dropped by 4% and 14.9% respectively. The introduction of the SPPCSPC_G grouped spatial pyramid pooling module improved the detection accuracy while compressing the network model; the third set of experiments used K-means++ to set the initial anchor box of the model while using the GELU function, which helps to improve the convergence speed and performance of the training process. Compared with the original YOLOv7 model, the detection accuracy is increased by 3.9%. The fourth set of experiments is the algorithm proposed in the article. Compared with the original YOLOv7 model, the detection accuracy reaches 72.3%, increased by 6.9%, and the amount of calculation and parameters decreased by 4% and 14.9% respectively, making it easier to deploy on edge terminal devices. The FPS reached 80, which is more suitable for real-time detection of road defects. The improved algorithm speeds up The convergence speed of the model has greatly improved the detection accuracy, achieving the purpose of detecting pavement diseases with a high recognition rate.

TABLE 6. Ablation experiment.

Group	K-means++	SPPCSPC_G	GELU	mAP@0.5%	FPS	Flops/G	Params/M
1	×	×	×	65.4	68	104.8	37.22
2	√	√	×	69.8	70	100.3	31.52
3	√	×	√	69.3	69	104.8	37.22
4	√	√	√	72.3	80	100.3	31.52

In order to test the actual effect of the above model, the following complex background images are used for detection. These images are all from outside the training set and verification set to ensure the reliability and validity of the image inspection. Figure 10 shows the above four sets of models for various types of diseases. detection effect.

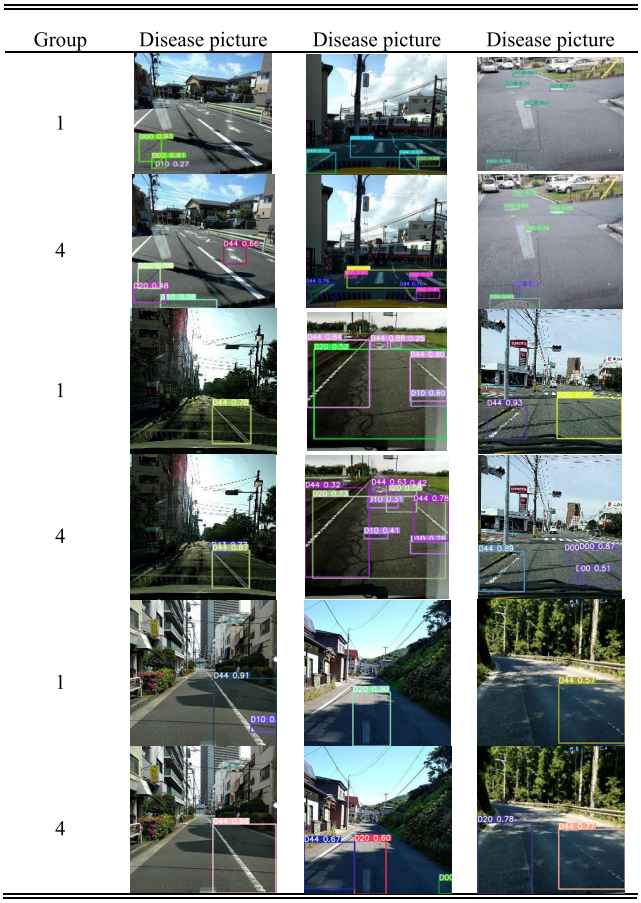


FIGURE 10. Comparison of detection effects of four groups of models.

It can be seen from the test results in Figure 10 that group 1 is the image effect of road defects detected by the unimproved algorithm, and group 4 is the image effect of road defects detected by the improved algorithm. It can be seen that the defect images identified by YOLOv7 have missed detections and wrong detections, and the detection effect of road defects in complex background is not good. The improved algorithm not only has high accuracy, but also has no missed detections or wrong detections. In this case, the reasoning speed is also much faster than the original YOLOv7, which proves that the improved YOLOv7 algorithm extracts richer semantic information and shows better performance.

5) OTHER PAVEMENT DEFECT DATA SET TESTING

In order to verify whether the algorithm in this paper has advantages on other data sets, experiments were conducted on the public data set RDD2022 [19] and the public data set global road damage detection challenge (GRDDC) [25]. To ensure the reliability of the experiments, they were all tested on the same equipment and under the same parameters. Table 7 is a comparison of experimental results of different data sets.

It can be seen from the experimental results that the average accuracy of the algorithm proposed in this article reached 67.8% on the RDD2022 data set, which is 3.5 percentage points higher than the original YOLOv7. On the GRDDC data set, the average accuracy of the algorithm proposed in this article reached 68.4%, which is 5.7% higher than the original YOLOv7. It can be seen that the method we proposed does not lack performance when extended to different image conditions.

TABLE 7. Comparison of experimental results on different data sets.

Data sat	Model	mAP@0.5%	FPS	Flops/G	Params/M
RDD2022	YOLOv7	64.3	43	117.5	35.61
RDD2022	Ours	67.8(+3.5)	59	100.3	31.52
GRDDC	YOLOv7	62.7	55	117.5	35.61
GRDDC	Ours	68.4(+5.7)	62	100.3	31.52

6) COMPARATIVE EXPERIMENTS OF DIFFERENT NETWORK MODELS

In order to further verify the effectiveness of the algorithm proposed in this article, under the premise of using the same training configuration and data set, other network models and the YOLOv7 network model improved in this article use the same parameters to train the data set.

TABLE 8. Comparative experiments of different network models.

Model	Backbone network	mAP@0.5%	FPS
Faster R-CNN	Resnext101	62.4	45
YOLOv3	CSPdarknet	59.7	52
YOLOv4	CSPdarknet	61	61
YOLOv5	CSPdarknet	63.5	65
YOLOv7		65.4	68
YOLOv8	CSPdarknet	67	74
Ours		72.3	80

As can be seen from Table 8, this paper has great advantages compared with mainstream target detection algorithms. This is because we replaced the original SPPCSPC with

SPPCSPC_G, which can reduce the amount of parameters and calculations while ensuring the detection effect; Secondly, this study uses the GELU function as the activation function of the REPConv convolution module in the YOLOv7 model, which solves the complexity of the model If it is too high, the model will have the problem of spatial insensitivity, thereby reducing model overfitting; then, reset the clustering prior framework to improve the randomness and locality of the cluster center selection process, thereby improving model detection accuracy. Therefore, the improved YOLOv7 algorithm further improves the detection accuracy, and the model size and parameter amount are also reduced. The optimized algorithm not only meets the requirements of real-time detection, but also can accurately identify complex background defects such as shadow occlusion and uneven lighting brightness. Reducing the amount of model parameters and calculations also makes it possible to deploy the model on edge terminal devices. Therefore, the algorithm in this paper is more suitable for defect detection in complex background than other algorithms.

V. COUNCLUSION

In order to improve the accuracy of intelligent detection of asphalt pavement diseases, this paper proposes a complex background defect detection algorithm based on improved YOLOv7. First, the K-means++ clustering algorithm is used to set the initial anchor box to obtain better anchor box parameters, thereby improving the defect detection accuracy for complex background such as shadow occlusion; then, the grouped spatial pyramid pooling module SPPCSPC_G is introduced into the head architecture to replace the original SPPCSPC module to reduce the calculation amount and parameter amount of the model and improve the fusion speed of image features; finally, the GELU activation function is used as the activation function of the REPConv convolution module in the YOLOv7 model, which effectively reduces the model overfitting, This improves the model's detection accuracy of road defects under complex backgrounds.

The test results show that the average accuracy of the improved detection algorithm for disease detection increased from 65.4% to 72.3%, an increase of 6.9%, the amount of calculation and parameters decreased by 4% and 14.9% respectively, and the FPS reached 80, an increase of 17%, which is more suitable for real-time detection of pavement defects, and no pavement defects are missed or wrongly detected. It can be seen that the improved YOLOv7 has better detection effect on complex background defects.

The algorithm in this article aims to improve the model's detection accuracy of complex background defects such as shadow occlusion and uneven light brightness, and can accurately identify road defects with shadow occlusion and uneven light brightness. It has good engineering application capabilities and can efficiently complete high-precision identification of complex background defects. The reduction of model parameters and calculations also provides a good solution for the deployment of edge terminal equipment, making it feasible to achieve damage ratings on complex background. In the follow-up work, we will conduct more in-depth research on the algorithm in this article in terms of model lightweighting and improving model detection accuracy, to further improve the practicality of the model, and thus meet the requirements of being deployed on edge terminal devices with limited computing resources.

REFERENCES

[1] Y. Fou and M. Shen, "The ministry of transport released the statistical bulletin on the development of the transportation industry in 2021," *Waterways Ports*, vol. 43, no. 3, p. 346, 2022.

[2] S. Xiaobo, "Research on pavement crack detection based on improved YOLOv4," *Henan Sci. Technol.*, vol. 41, no. 18, pp. 62–67, 2022, doi: 10.19968/j.cnki.hnkj.1003-5168.2022.8.012.

[3] Z. Zhihua, D. Yanxue, and Z. Xinxiu, "Asphalt pavement disease extraction and classification method based on improved SegNet," *Traffic Inf. Saf.*, vol. 40, no. 3, pp. 127–135, 2022.

[4] L. Erwei, "Crack identification technology of Binhai Bridge based on machine vision," *Construct. Saf.*, vol. 38, no. 2, pp. 33–37, 2023.

[5] L. Xing, Z. Jiang, W. Weikang, Y. Shiji, and L. Xin, "PeleeNet_YOLOv3 surface crack identification using lightweight model," *J. Harbin Inst. Technol.*, vol. 55, no. 4, pp. 81–89, 2023.

[6] H. Tiejun, and L. Hua'en, "Pavement disease detection model based on improved YOLOv5," *J. Civil Eng.*, vol. 57, no. 2, pp. 96–106, 2024.

[7] C. Jianyu, Z. Chunlong, and W. Shenghuai, "Research on improving YOLOv5 rapid detection method of road defects," *Electron. Meas. Technol.*, vol. 46, no. 10, pp. 129–135, 2023, doi: 10.19651/j.cnki.emt.2211195.

[8] Z. Hao and N. Xiaowei, "Research on pavement defect detection algorithm based on improved YOLOv5," *Inf. Technol. Informatization*, no. 1, pp. 50–53, 2023.

[9] H. Yang, Y. Liu, S. Wang, H. Qu, N. Li, J. Wu, Y. Yan, H. Zhang, J. Wang, and J. Qiu, "Improved apple fruit target recognition method based on YOLOv7 model," *Agriculture*, vol. 13, no. 7, p. 1278, Jun. 2023.

[10] Y. Zhang, Y. Sun, Z. Wang, and Y. Jiang, "YOLOv7-RAR for urban vehicle detection," *Sensors*, vol. 23, no. 4, p. 1801, Feb. 2023.

[11] K. Liu, Q. Sun, D. Sun, L. Peng, M. Yang, and N. Wang, "Underwater target detection based on improved YOLOv7," *J. Mar. Sci. Eng.*, vol. 11, no. 3, p. 677, Mar. 2023.

[12] W. Zhigao and C. Ming, "Lightweight detection method of microalgae based on improved YOLO v7," *J. Dalian Ocean Univ.*, vol. 38, no. 1, pp. 129–139, 2023.

[13] P. Huang, S. Wang, J. Chen, W. Li, and X. Peng, "Lightweight model for pavement defect detection based on improved YOLOv7," *Sensors*, vol. 23, no. 16, p. 7112, Aug. 2023.

[14] N. Changshuang et al., "Improving asphalt pavement disease detection of YOLOv7," *Comput. Eng. Appl.*, vol. 59, no. 13, pp. 305–316, 2023.

[15] P. Liu and H. Yin, "YOLOv7-peach: An algorithm for immature small yellow peaches detection in complex natural environments," *Sensors*, vol. 23, no. 11, p. 5096, May 2023.

[16] J. Chen, B. Ma, C. Ji, J. Zhang, Q. Feng, X. Liu, and Y. Li, "Apple inflorescence recognition of phenology stage in complex background based on improved YOLOv7," *Comput. Electron. Agricult.*, vol. 211, Aug. 2023, Art. no. 108048.

[17] Z. Zhang, "Traffic sign detection algorithm based on improved YOLOv7," *Proc. SPIE*, vol. 12707, pp. 1258–1266, Jun. 2023.

[18] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 7464–7475.

[19] H. Maeda, T. Kashiyama, Y. Sekimoto, T. Seto, and H. Omata, "Generative adversarial network for road damage detection," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 36, no. 1, pp. 47–60, Jan. 2021.

[20] Ö. Kaya, M. Y. Çodur, and E. Mustafaraj, "Automatic detection of pedestrian crosswalk with faster R-CNN and YOLOv7," *Buildings*, vol. 13, no. 4, p. 1070, Apr. 2023.

[21] Y. Que, Y. Dai, X. Ji, A. Kwan Leung, Z. Chen, Z. Jiang, and Y. Tang, "Automatic classification of asphalt pavement cracks using a novel integrated generative adversarial networks and improved VGG model," *Eng. Struct.*, vol. 277, Feb. 2023, Art. no. 115406.
[22] C. Wu, M. Ye, J. Zhang, and Y. Ma, "YOLO-LWNet: A lightweight road damage object detection network for mobile terminal devices," *Sensors*, vol. 23, no. 6, p. 3268, Mar. 2023.
[23] Z. Yang, C. Ni, L. Li, W. Luo, and Y. Qin, "Three-stage pavement crack localization and segmentation algorithm based on digital image processing and deep learning techniques," *Sensors*, vol. 22, no. 21, p. 8459, Nov. 2022.
[24] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," 2016, *arXiv:1606.08415*.
[25] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, H. Omata, T. Kashiyama, and Y. Sekimoto, "Global road damage detection: State-of-the-art solutions," in *Proc. IEEE Int. Conf. Big Data*, Dec. 2020, pp. 5533–5539.

ZOU CHUNLONG was born in Xiangyang, Hubei, in 1988. He received the bachelor's degree from Hubei University of Arts and Sciences, in 2011, and the master's degree from Wuhan University of Science and Technology, in 2014. In 2016, he was a Teacher at Hubei University of Automotive Technology.

HUANG PEILE was born in January 2000. He received the bachelor's degree from Weifang University of Science and Technology, in 2022. He is currently a 22nd-level graduate student with the Hubei University of Automotive Industry. His main research directions are artificial intelligence and machine vision technology.

WANG SHENGHUAI received the Ph.D. degree from Huazhong University of Science and Technology, in 2009. He is currently a Professor and a Ph.D. Supervisor with the Department of Mechanical Engineering, Hubei University of Automotive Technology. His research interest includes precision measurement technology.

WANG CHEN was born in Shiyan, Hubei, China, in 1983. He received the B.S. degree from Hubei University of Automotive Technology, in 2006, the M.S. degree from Wuhan University of Science and Technology, in 2009, and the Ph.D. degree in mechanical engineering from Shanghai University, in 2019. Since 2018, he has been an Associate Professor with the Department of Mechanical Engineering, Hubei University of Automotive Technology.

WANG HONGXIA received the Ph.D. degree from Chongqing University. She is currently the Deputy Dean with the School of Mechanical Engineering, Hubei University of Automotive Industry. His research directions are mechanical design theory, vibration, and noise analysis.